Homogenization in Multi-Agent Systems

Prakhar Ganesh[1,2] †, Kyra Wilson[3] †, Luca Zappella[4], Barry-John Theobald[4], Nicholas Apostoloff[4], Lucas Monteiro Paes[4] *, Nivedha Sivakumar[4] *

Special thanks to: Katherine Metcalf[4], Jerremy Holland[4], Natalie Mackraz[4]

[1] McGill University
[2] Mila - Quebec AI Institute
[3] University of Washington
[4] Apple

† work done during internship at Apple
*equal supervisory contribution

About Me



Prakhar Ganesh (he/him)

PhD student in Computer Science
at McGill University / Mila

Research in Multiplicity & Homogenization in AI
with an emphasis on Fairness

Background and Motivation



Homogenization in AI

Homogenization in AI

"[...] outcome homogenization: the extent to which particular individuals or groups experience negative outcomes from all decision-makers." - Bommasani et al., (2022)

Bommasani, Rishi, et al. "Picking on the same person: Does algorithmic monoculture lead to outcome homogenization?." Advances in neural information processing systems 35 (2022): 3663-3678.

Homogenization in AI

"[...] outcome homogenization: the extent to which particular individuals or groups experience negative outcomes from all decision-makers." - Bommasani et al., (2022)

"[...] outcome homogenization, where independent decision-makers in a given domain reach the same conclusion" - Gur-Arieh and Lee, (2025)

Bommasani, Rishi, et al. "Picking on the same person: Does algorithmic monoculture lead to outcome homogenization?." Advances in neural information processing systems 35 (2022): 3663-3678.
Gur-Arieh, Shira, and Christina Lee. "Consistently arbitrary or arbitrarily consistent: Navigating the tensions between homogenization and multiplicity in algorithmic decision-making." Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency. 2025.

Homogenization in AI

"[...] outcome homogenization: the extent to which particular individuals or groups experience negative outcomes from all decision-makers." – Bommasani et al., (2022)

"[...] outcome homogenization, where independent decision-makers in a given domain reach the same conclusion" – Gur-Arieh and Lee, (2025)

"[...] homogenization in LLMs, defined simply as models failing to generate diverse outputs." – Chehbouni et al., (2026)

Bommasani, Rishi, et al. "Picking on the same person: Does algorithmic monoculture lead to outcome homogenization?." Advances in neural information processing systems 35 (2022): 3663-3678.
Gur-Arieh, Shira, and Christina Lee. "Consistently arbitrary or arbitrarily consistent: Navigating the tensions between homogenization and multiplicity in algorithmic decision-making." Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency. 2025.
Chehbouni, Khaoula, et al. "From Infrastructure to Interface, the AI Value Chain Drives LLM Homogenization" Advances in Neural Information Processing Systems (2026).

When is Homogenization Harmful?

- **Factual vs Normative Decisions:** Whether there exists a single ground truth for the target variable
- **Single-Shot vs Multi-Shot Settings**: Whether individuals may receive the resource or opportunity from another decision-maker at a later time
- **Low vs High Aleatoric Uncertainty**: Whether there is variability in the target variable for the same inputs
- **Public vs Private Decisions**: Whether decision-makers are public entities or private entities

Gur-Arieh, Shira, and Christina Lee. "Consistently arbitrary or arbitrarily consistent: Navigating the tensions between homogenization and multiplicity in algorithmic decision-making." Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency. 2025.

Homogenization in AI

The Algorithmic Leviathan: Arbitrariness, Fairness, and Opportunity in Algorithmic Decision-Making Systems

Kathleen Creel*,[1] and Deborah Hellman[2]

Algorithmic Monocultures in Hiring

RISHI BOMMASANI*, Stanford University, USA
SARAH H. BANA*, Chapman University, USA
KATHLEEN A. CREEL*, Northeastern University, USA
DAN JURAFSKY, Stanford University, USA
PERCY LIANG, Stanford University, USA

Generative Monoculture in Large Language Models

Fan Wu[1], Emily Black[2], Varun Chandrasekaran[1]
[1] University of Illinois Urbana-Champaign [2] Barnard College

Artificial Hivemind: The Open-Ended Homogeneity of Language Models (and Beyond)

Liwei Jiang♠ Yuanjun Chai♠ Margaret Li♠ Mickel Liu♠ Raymond Fok♠
Nouha Dziri★ Yulia Tsvetkov♠ Maarten Sap◇ Alon Albalak♣ Yejin Choi♡
♠University of Washington ◇ Carnegie Mellon University
★Allen Institute for Artificial Intelligence ♣Lila Sciences ♡Stanford University

Multi-Agent Systems

Multi-Agent Systems orchestrate interactions among agents by dividing complex tasks. Through critique, debate, and refinement, these interactions produce outcomes that can exceed what isolated agents achieve on their own.



Image taken from
https://exploringartificialintelligence.substack.com/p/architecture-for-multi-agent-systems

Interaction Dynamics in Multi-Agent Systems

From Belief Entrenchment to Robust Reasoning in LLM Agents

Jihwan Oh[1*] **Minchan Jeong**[1*] **Jongwoo Ko**[2†] **Se-Young Yun** [1†]
[1]KAIST AI [2]Microsoft
{ericoh929, mcjeong}@kaist.ac.kr

ALIGNED AGENTS, BIASED SWARM: MEASURING BIAS AMPLIFICATION IN MULTI-AGENT SYSTEMS

Keyu Li[1,2], **Jin Gao**[1], **and Dequan Wang**[1,2†]
[1]Shanghai Jiao Tong University [2]Shanghai Innovation Institute

Towards Simulating Social Influence Dynamics with LLM-based Multi-agents

Hsien-Tsung Lin
Department of Information Management
National Sun Yat-Sen University
Kaohsiung, Taiwan
aidenlin@g-mail.nsysu.edu.tw

Pei-Cing Huang
Department of Information Management
National Sun Yat-Sen University
Kaohsiung, Taiwan
pcpeicing@gmail.com

Chan-Tung Ku
Department of Information Management
National Sun Yat-Sen University
Kaohsiung, Taiwan
kuchantung@gmail.com

Chan Hsu
Department of Information Management
National Sun Yat-Sen University
Kaohsiung, Taiwan
chanshsu@gmail.com

Pei-Xuan Shieh
Department of Information Management
National Sun Yat-Sen University
Kaohsiung, Taiwan
iamshiehpay@gmail.com

Yihuang Kang
Department of Information Management
National Sun Yat-Sen University
Kaohsiung, Taiwan
ykang@mis.nsysu.edu.tw

Emergent social conventions and collective bias in LLM populations

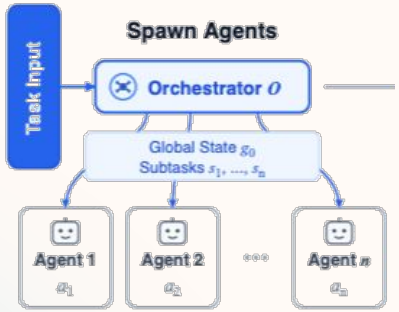
Ariel Flint Ashery[1], **Luca Maria Aiello**[2,3], **Andrea Baronchelli**[1,4*]

Definitions and Metrics

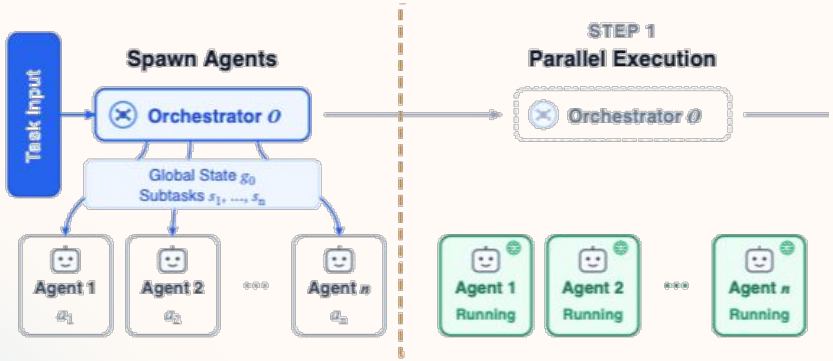


Orchestrator-driven Multi-Agent System

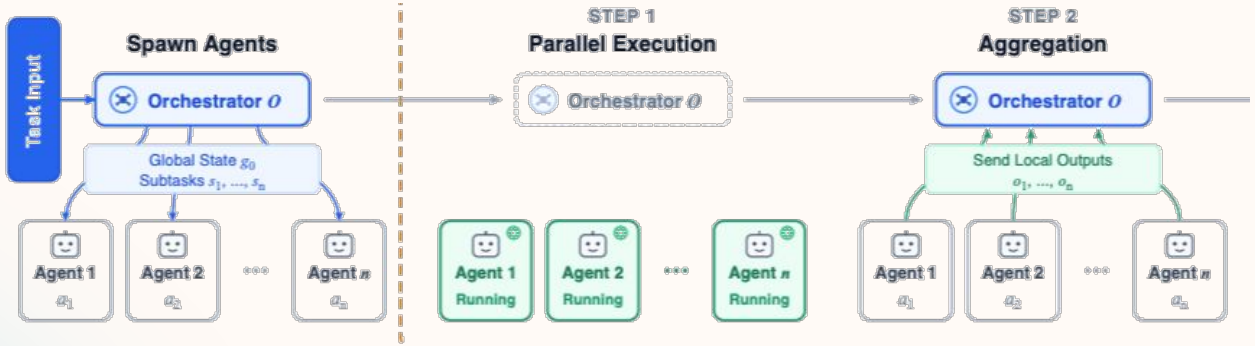
Orchestrator-driven Multi-Agent System



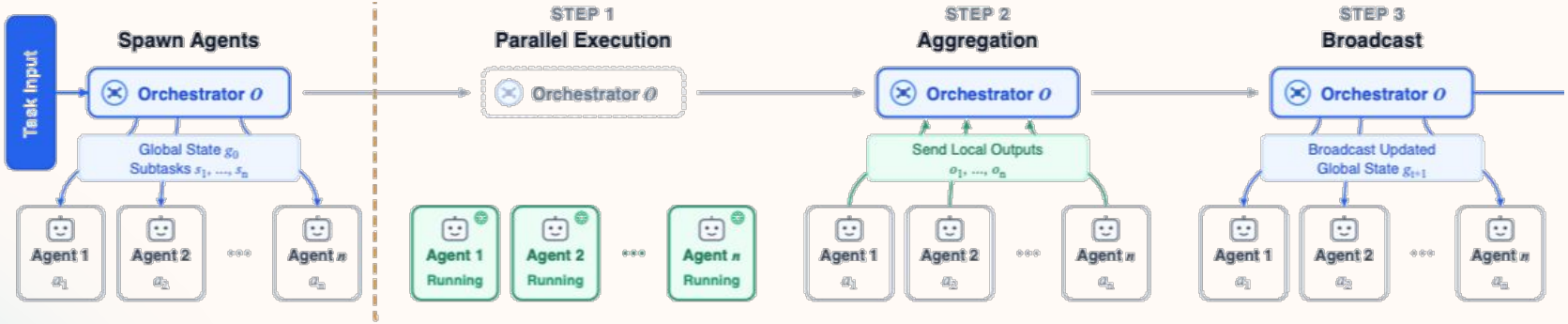
Orchestrator-driven Multi-Agent System



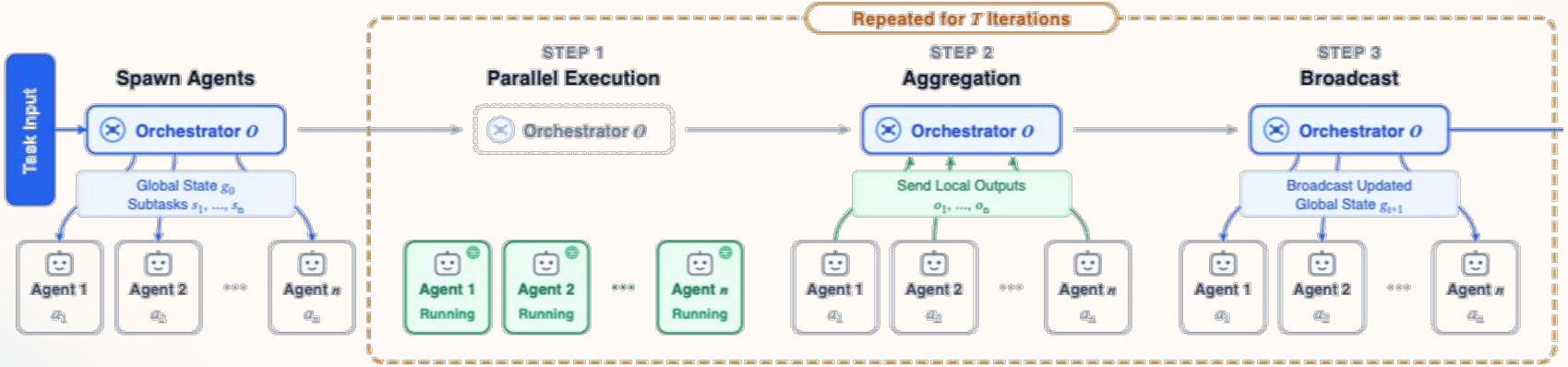
Orchestrator-driven Multi-Agent System



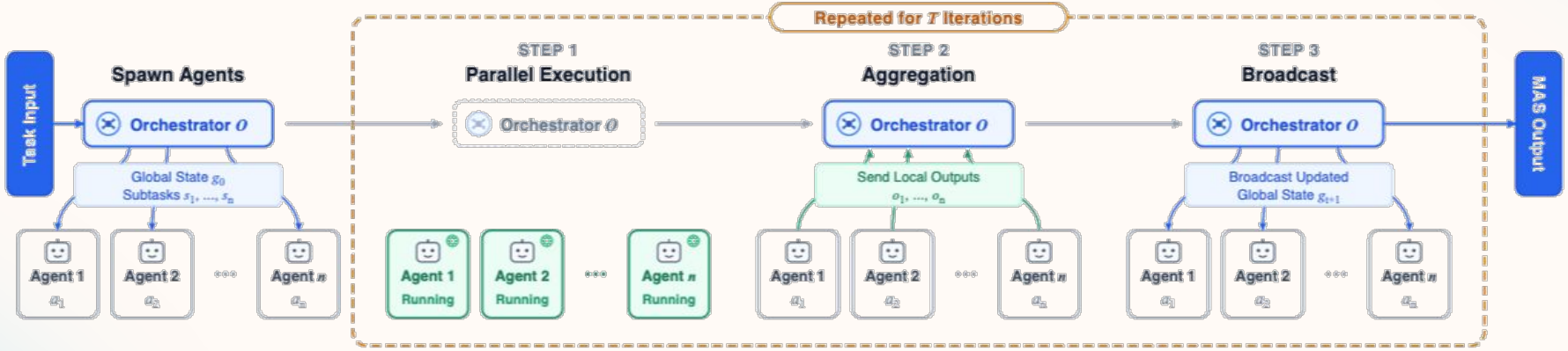
Orchestrator-driven Multi-Agent System



Orchestrator-driven Multi-Agent System



Orchestrator-driven Multi-Agent System

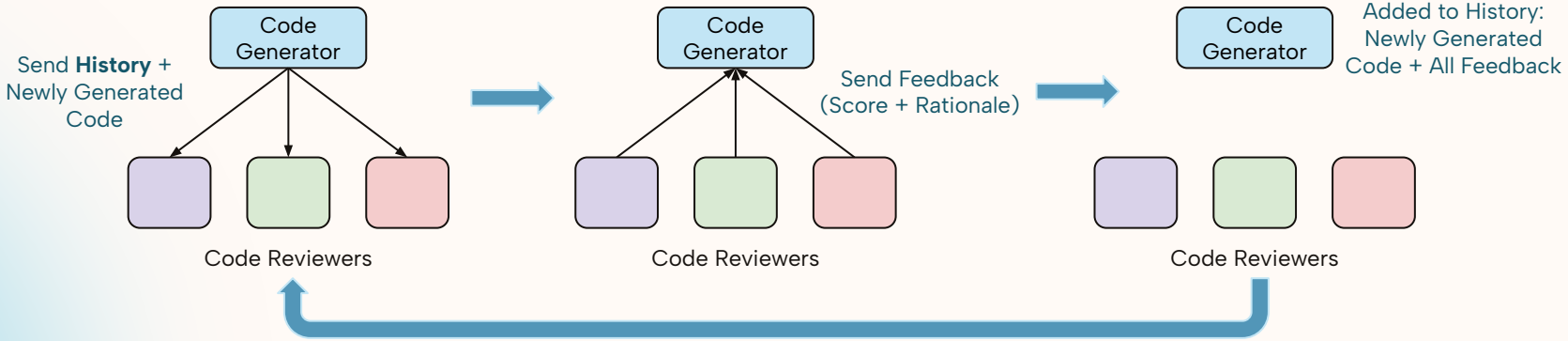


Setup Overview

	Code Generation	Hiring	Peer Review
Objective	Generate code & evaluate deployability	Résumé fit for job description	Review paper & give recommendation
Workers $\mathcal{A}$	Code Reviewers: Security, Optimality, Correctness, Readability	Evaluators: Subject Matter, Experience, Impact, Operational Readiness	Paper Reviewers: Pragmatist, Theorist, Empiricist
Score $r_i \in \mathcal{R}$	Deployability: $[1, 5]$ *(Not Ready → Ready)*	Résumé Fit: $[1, 5]$ *(Strong No-Hire → Strong Hire)*	Acceptance: $[1, 10]$ *(Strong Reject → Strong Accept)*
Rationale c_i	Code improvement suggestions	Fit rationale & qualification assessment	Structured review (with strengths, weaknesses, and questions)
Orchestrator O	Code generator, which incorporates worker feedback in later rounds	None, just direct concatenation	Meta-reviewer to summarize discussion & Author to provide rebuttal

Setup Overview

	Code Generation	Hiring	Peer Review
Objective	Generate code & evaluate deployability	Résumé fit for job description	Review paper & give recommendation
Workers $\mathcal{A}$	Code Reviewers: Security, Optimality, Correctness, Readability	Evaluators: Subject Matter, Experience, Impact, Operational Readiness	Paper Reviewers: Pragmatist, Theorist, Empiricist
Score $r_i \in \mathcal{R}$	Deployability: $[1, 5]$ *(Not Ready → Ready)*	Résumé Fit: $[1, 5]$ *(Strong No-Hire → Strong Hire)*	Acceptance: $[1, 10]$ *(Strong Reject → Strong Accept)*
Rationale c_i	Code improvement suggestions	Fit rationale & qualification assessment	Structured review (with strengths, weaknesses, and questions)
Orchestrator O	Code generator, which incorporates worker feedback in later rounds	None, just direct concatenation	Meta-reviewer to summarize discussion & Author to provide rebuttal



Setup Overview

	Code Generation	Hiring	Peer Review
Objective	Generate code & evaluate deployability	Résumé fit for job description	Review paper & give recommendation
Workers $\mathcal{A}$	Code Reviewers: Security, Optimality, Correctness, Readability	Evaluators: Subject Matter, Experience, Impact, Operational Readiness	Paper Reviewers: Pragmatist, Theorist, Empiricist
Score $r_i \in \mathcal{R}$	Deployability: $[1, 5]$ *(Not Ready → Ready)*	Résumé Fit: $[1, 5]$ *(Strong No-Hire → Strong Hire)*	Acceptance: $[1, 10]$ *(Strong Reject → Strong Accept)*
Rationale c_i	Code improvement suggestions	Fit rationale & qualification assessment	Structured review (with strengths, weaknesses, and questions)
Orchestrator O	Code generator, which incorporates worker feedback in later rounds	None, just direct concatenation	Meta-reviewer to summarize discussion & Author to provide rebuttal

Setup Overview

	Code Generation	Hiring	Peer Review
Objective	Generate code & evaluate deployability	Résumé fit for job description	Review paper & give recommendation
Workers $\mathcal{A}$	Code Reviewers: Security, Optimality, Correctness, Readability	Evaluators: Subject Matter, Experience, Impact, Operational Readiness	Paper Reviewers: Pragmatist, Theorist, Empiricist
Score $r_i \in \mathcal{R}$	Deployability: $[1, 5]$ *(Not Ready → Ready)*	Résumé Fit: $[1, 5]$ *(Strong No-Hire → Strong Hire)*	Acceptance: $[1, 10]$ *(Strong Reject → Strong Accept)*
Rationale c_i	Code improvement suggestions	Fit rationale & qualification assessment	Structured review (with strengths, weaknesses, and questions)
Orchestrator O	Code generator, which incorporates worker feedback in later rounds	None, just direct concatenation	Meta-reviewer to summarize discussion & Author to provide rebuttal

Metrics for Homogenization

Metrics for Homogenization

Conformity: Tendency of agents to conform, instead of holding their views.

Definition 3.2 (Conformity). *An orchestrator-driven MAS is said to exhibit conformity between rounds t and $t+1$ if the variance of agents' scores decreases across successive rounds, i.e.,*

$$\mathrm{Var}\left(r_1^{(t+1)}, \ldots, r_n^{(t+1)}\right) < \mathrm{Var}\left(r_1^{(t)}, \ldots, r_n^{(t)}\right),$$

where $r_i^{(t)}$ denotes the score produced by agent A_i at round t and $\mathrm{Var}(\cdot)$ denotes the variance of its arguments.

Metrics for Homogenization

Conformity: Tendency of agents to conform, instead of holding their views.

Polarization: Tendency of agents to get more confident and extreme.

Definition 3.3 (Polarization). *An orchestrator-driven MAS is said to exhibit polarization between rounds t and $t+1$ if the mean score across agents moves closer to extremes. Specifically, let $\bar{r}^{(t)} = \frac{1}{n}\sum_{i=1}^{n} r_i^{(t)}$ denote the mean score of agents at round t, polarization occurs when*

$$\min\left(\left|\bar{r}^{(t+1)} - r_{\min}\right|, \left|\bar{r}^{(t+1)} - r_{\max}\right|\right) < \min\left(\left|\bar{r}^{(t)} - r_{\min}\right|, \left|\bar{r}^{(t)} - r_{\max}\right|\right).$$

Metrics for Homogenization

Conformity: Tendency of agents to conform, instead of holding their views.

Polarization: Tendency of agents to get more confident and extreme.

Inertia: Tendency of agents to resist against changes with more interactions.

Definition 3.4 (Inertia). *An orchestrator-driven MAS is said to exhibit inertia between rounds t and $t+1$ if the rate of change in the mean score between rounds decreases across successive rounds. Specifically, let $\bar{r}^{(t)} = \frac{1}{n}\sum_{i=1}^{n} r_i^{(t)}$ denote the mean score of agents at round t, inertia occurs when*

$$\left|\bar{r}^{(t+2)} - \bar{r}^{(t+1)}\right| < \left|\bar{r}^{(t+1)} - \bar{r}^{(t)}\right|.$$

Metrics for Homogenization



Metrics for Homogenization



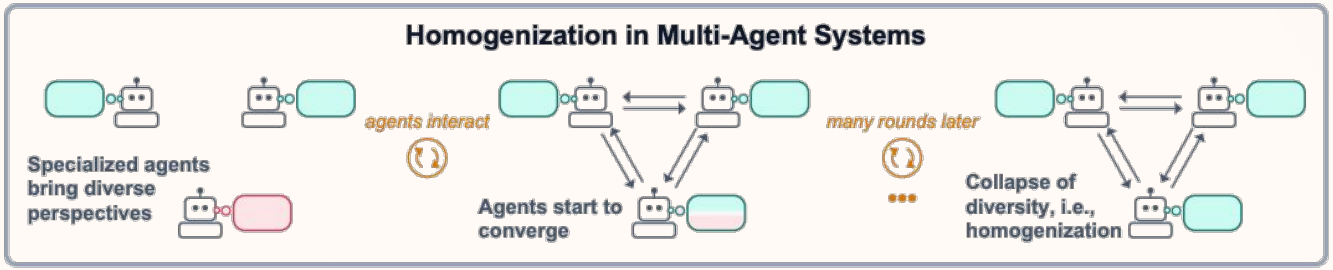
Metrics for Homogenization



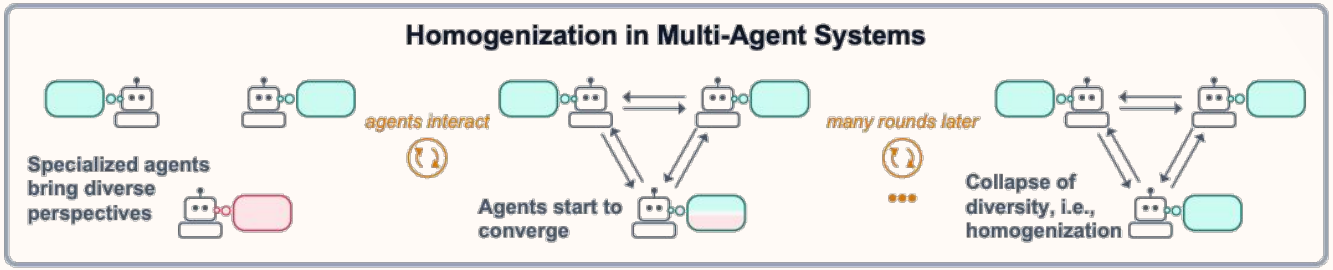
Our Expectations

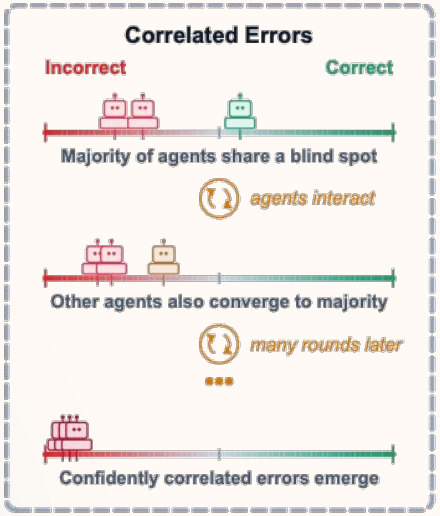


Types of Harms

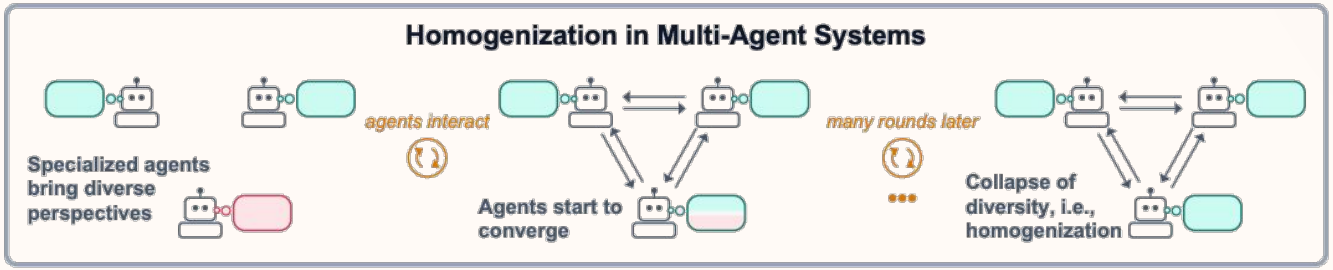


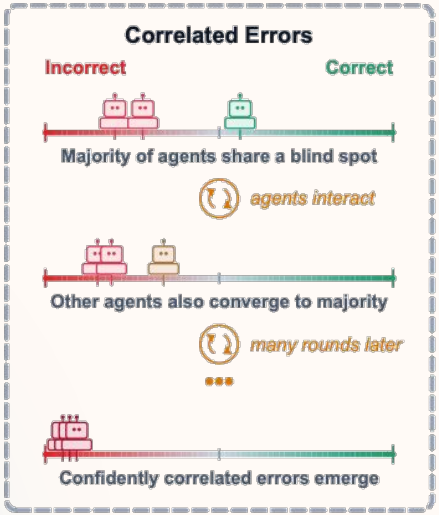
Types of Harms

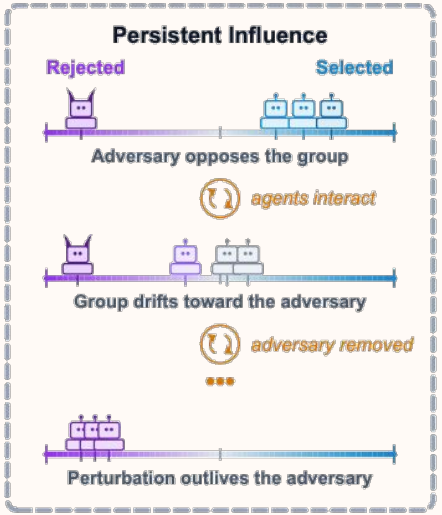




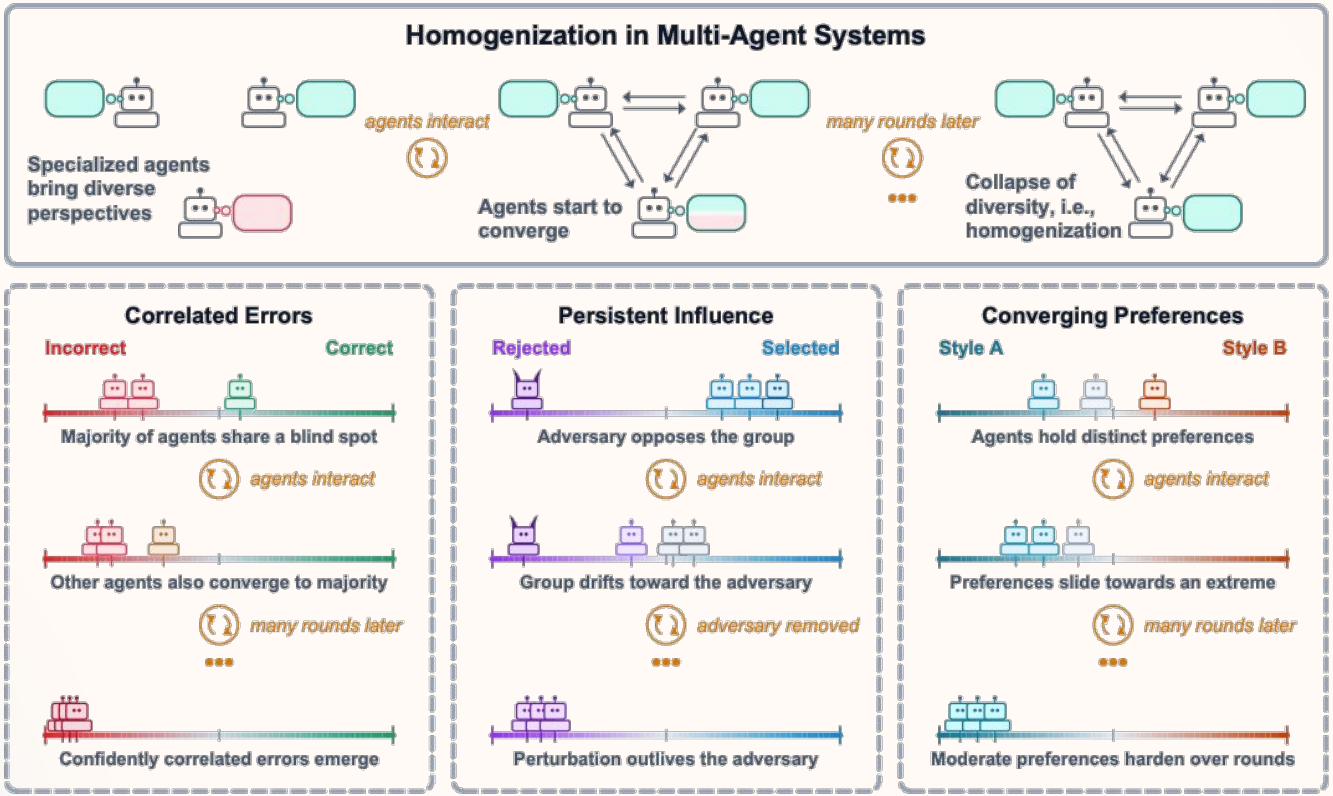
Types of Harms







Types of Harms

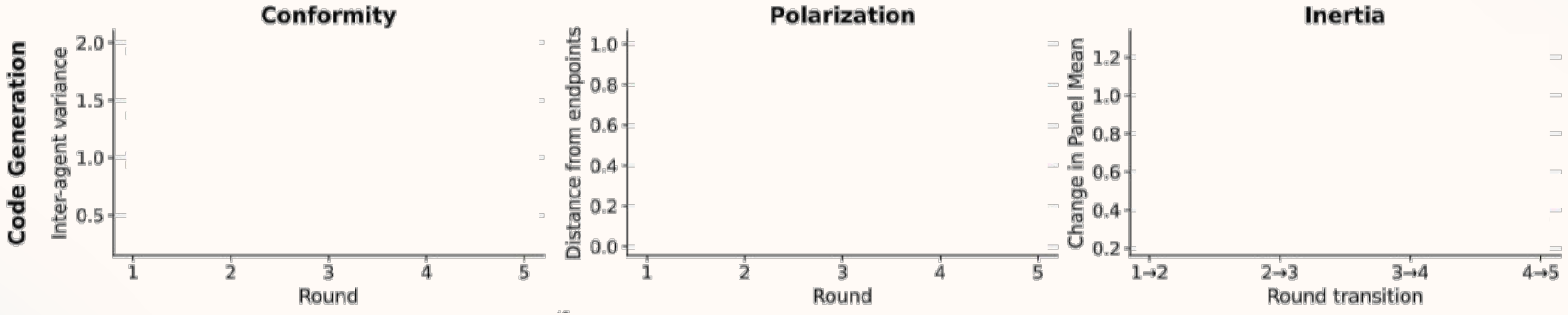


Homogenization Trends and Downstream Harms

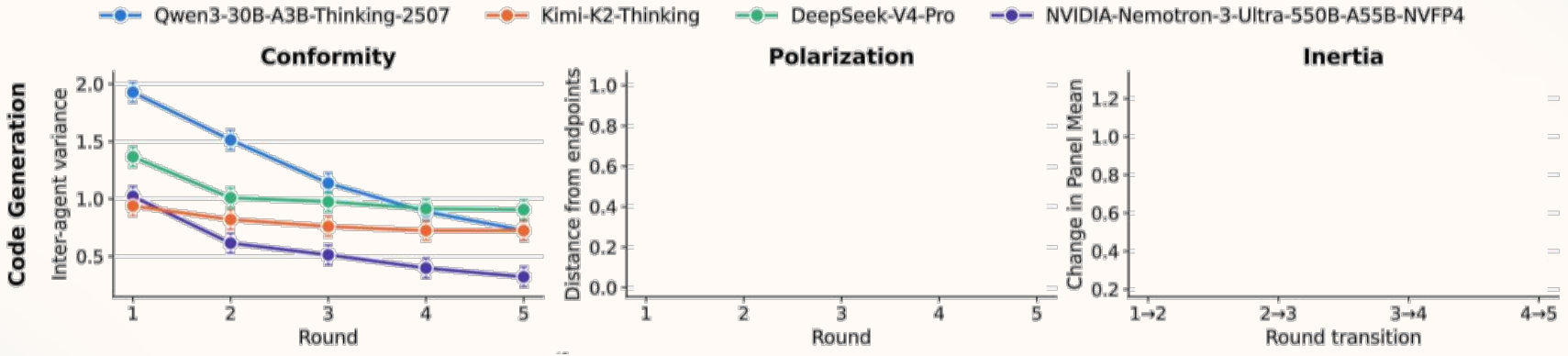


Multi-Agent Systems for Coding

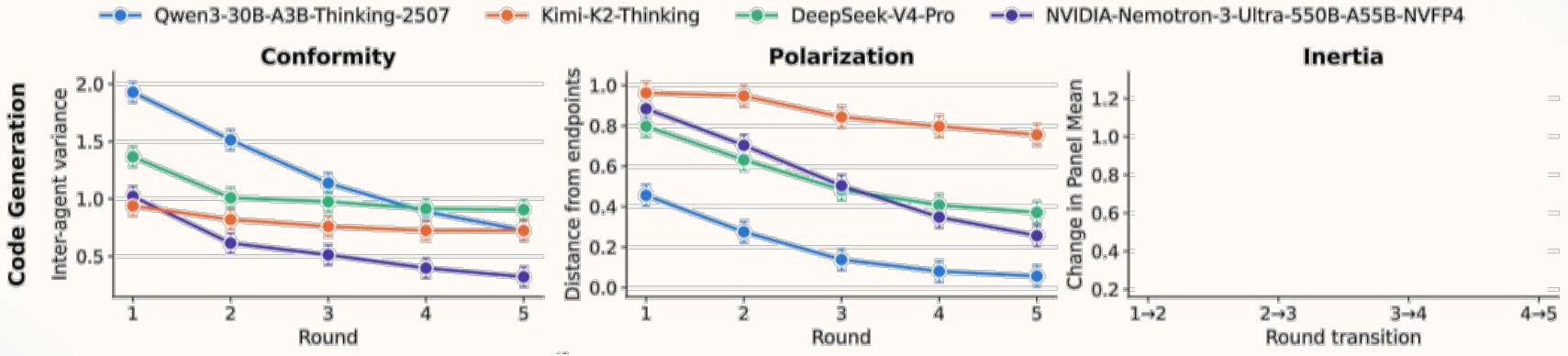
Homogenization Trends in Coding



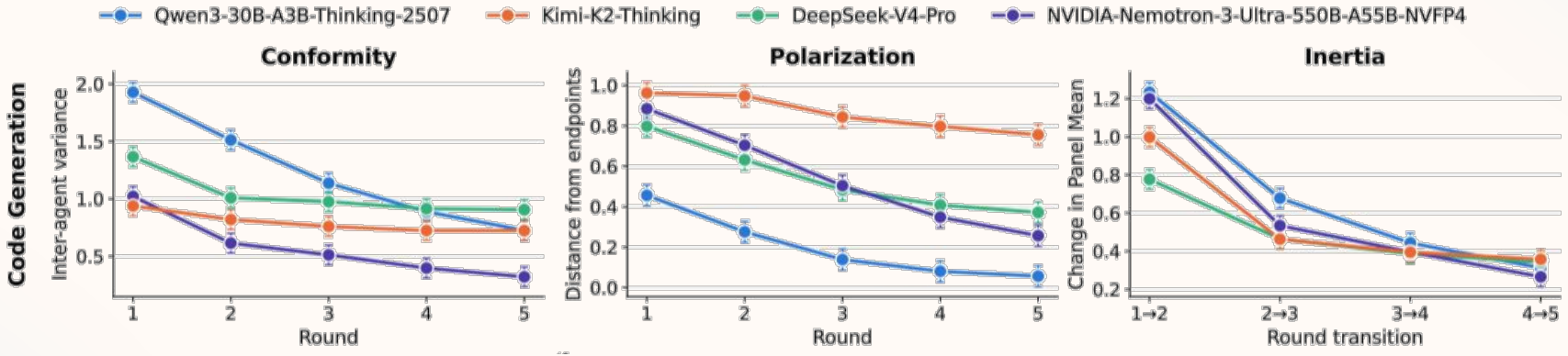
Homogenization Trends in Coding



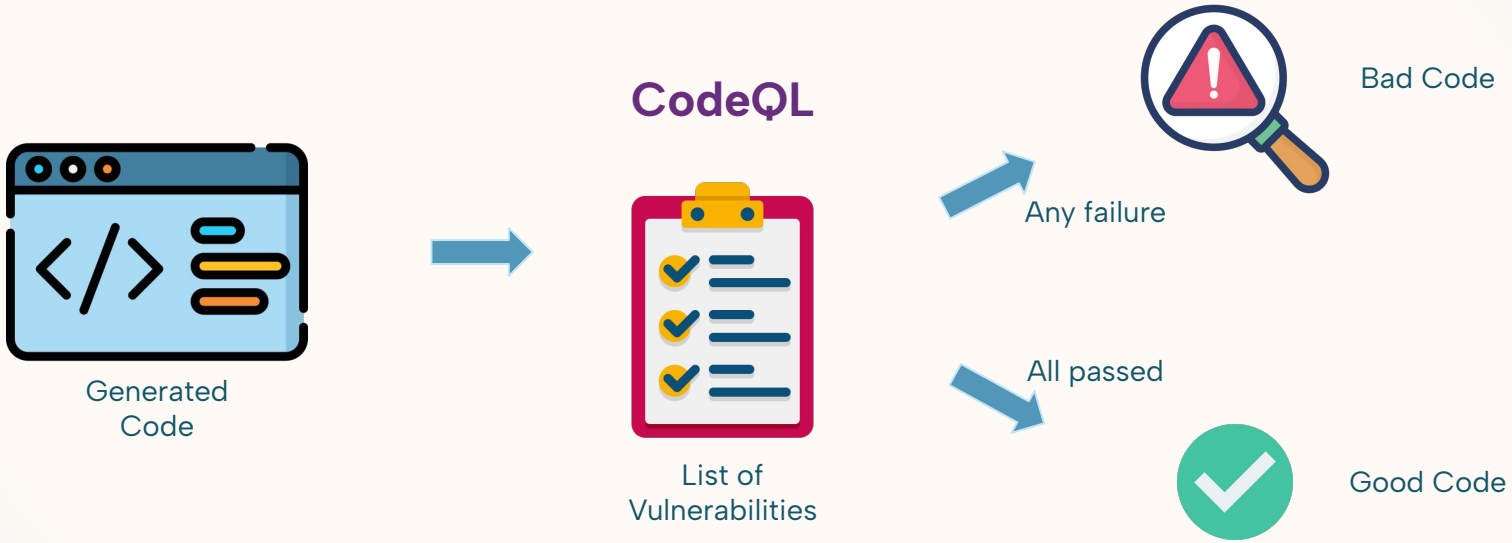
Homogenization Trends in Coding



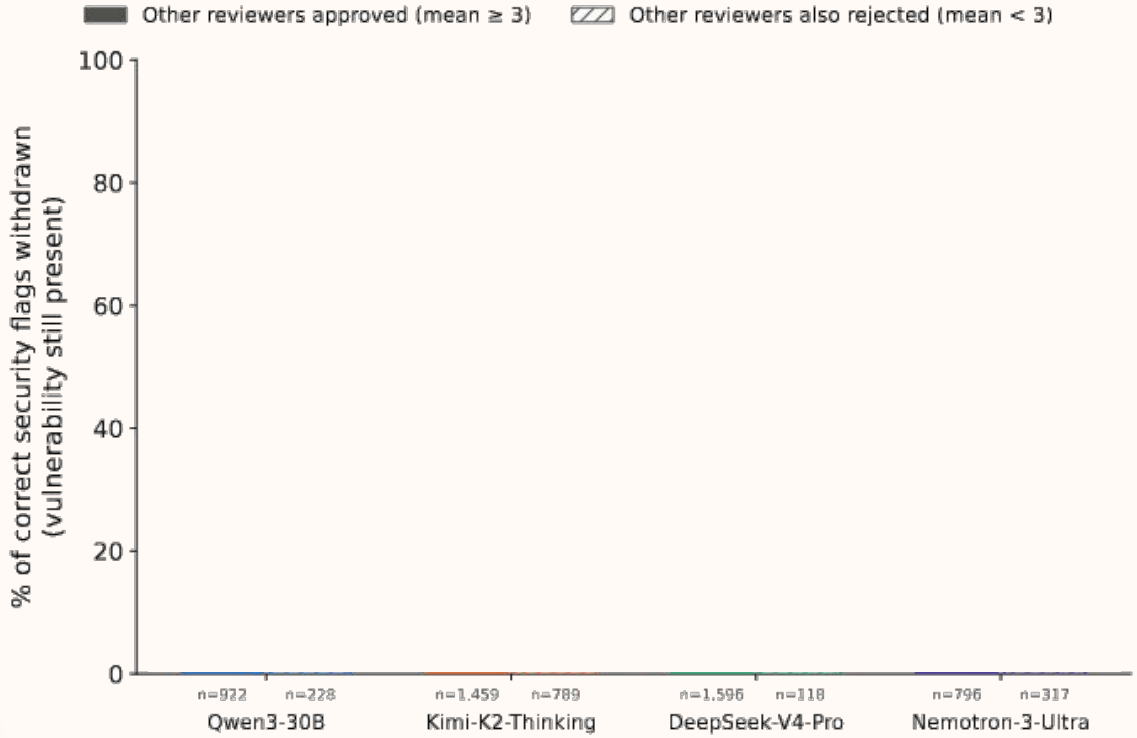
Homogenization Trends in Coding



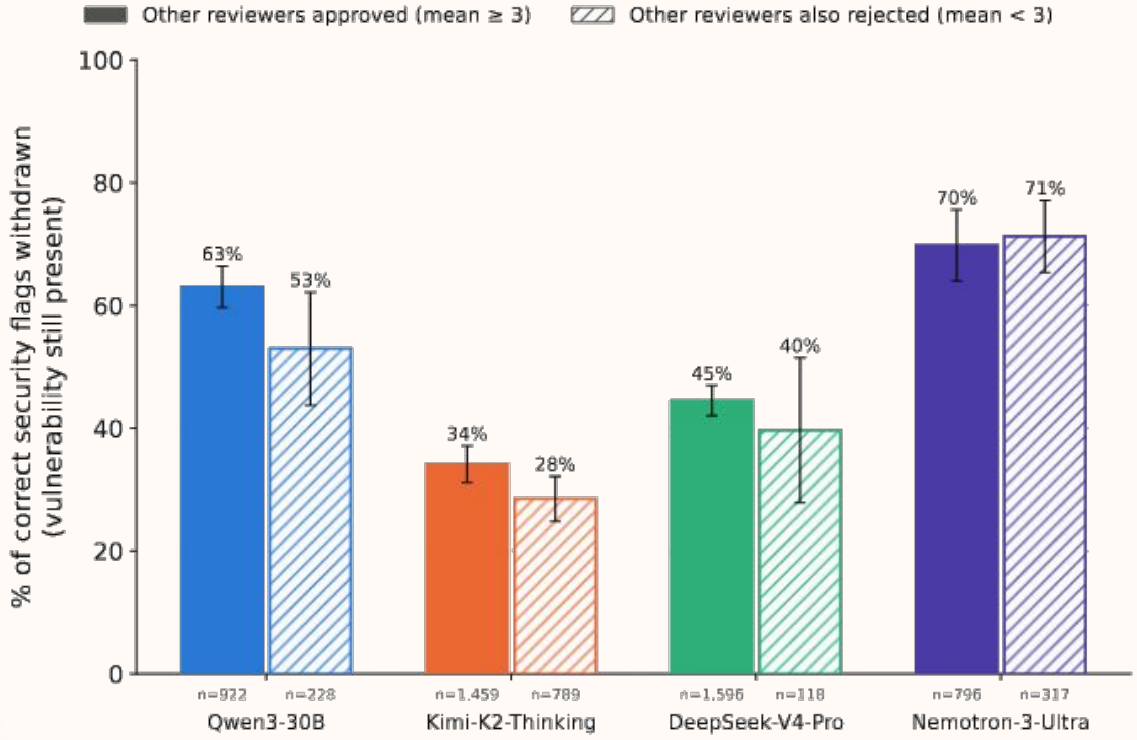
Evaluation of Code for Security Vulnerabilities



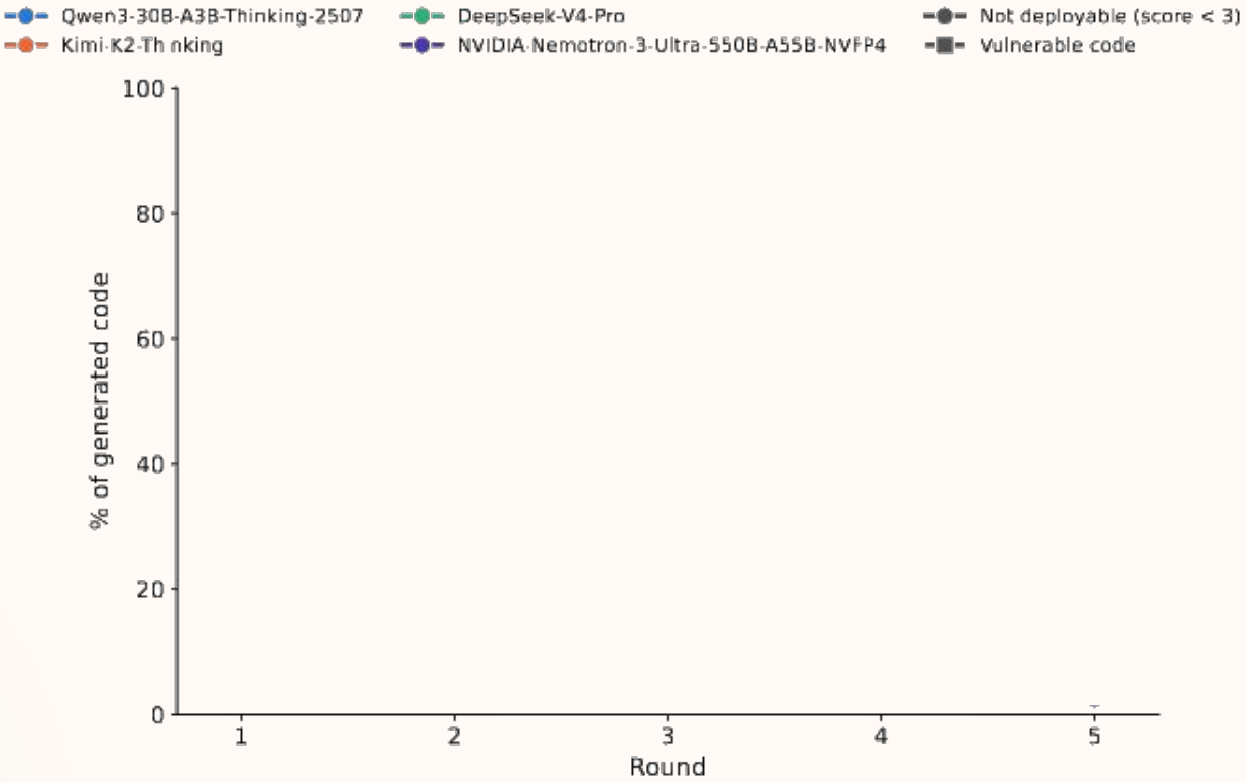
Impact of Conformity on Security Agent



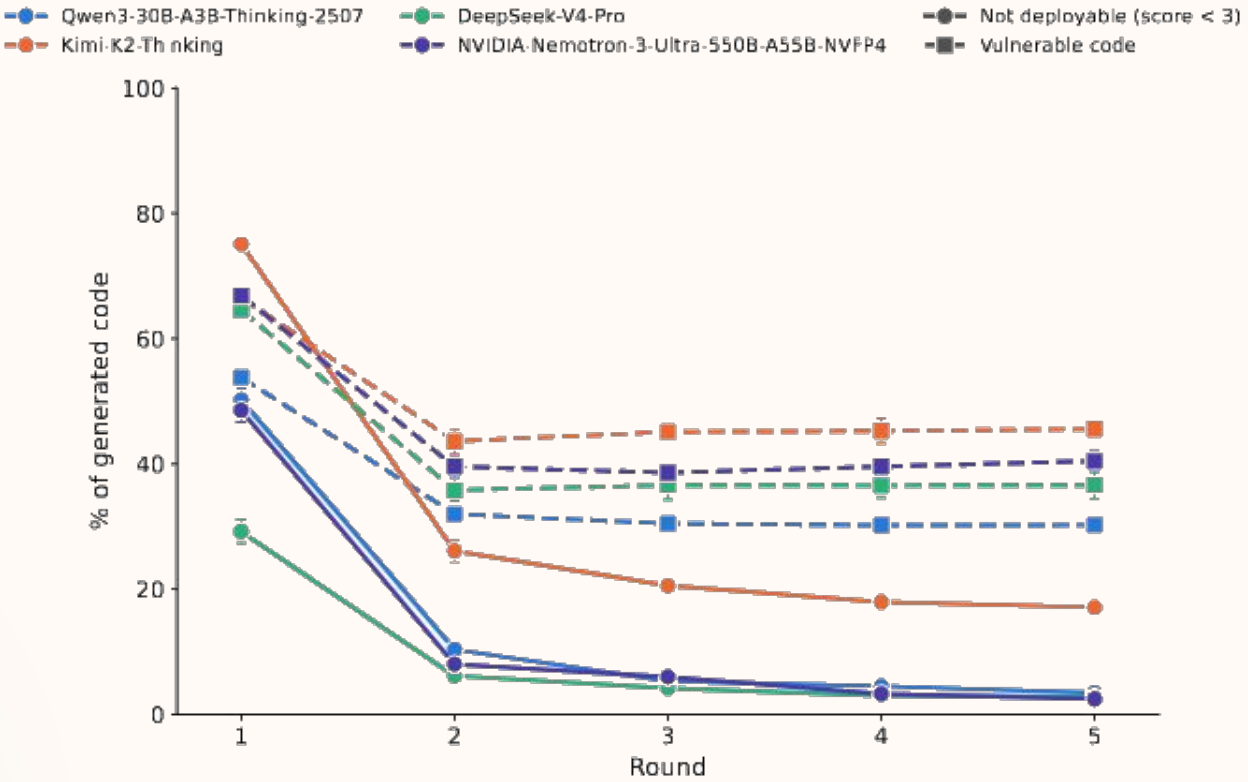
Impact of Conformity on Security Agent



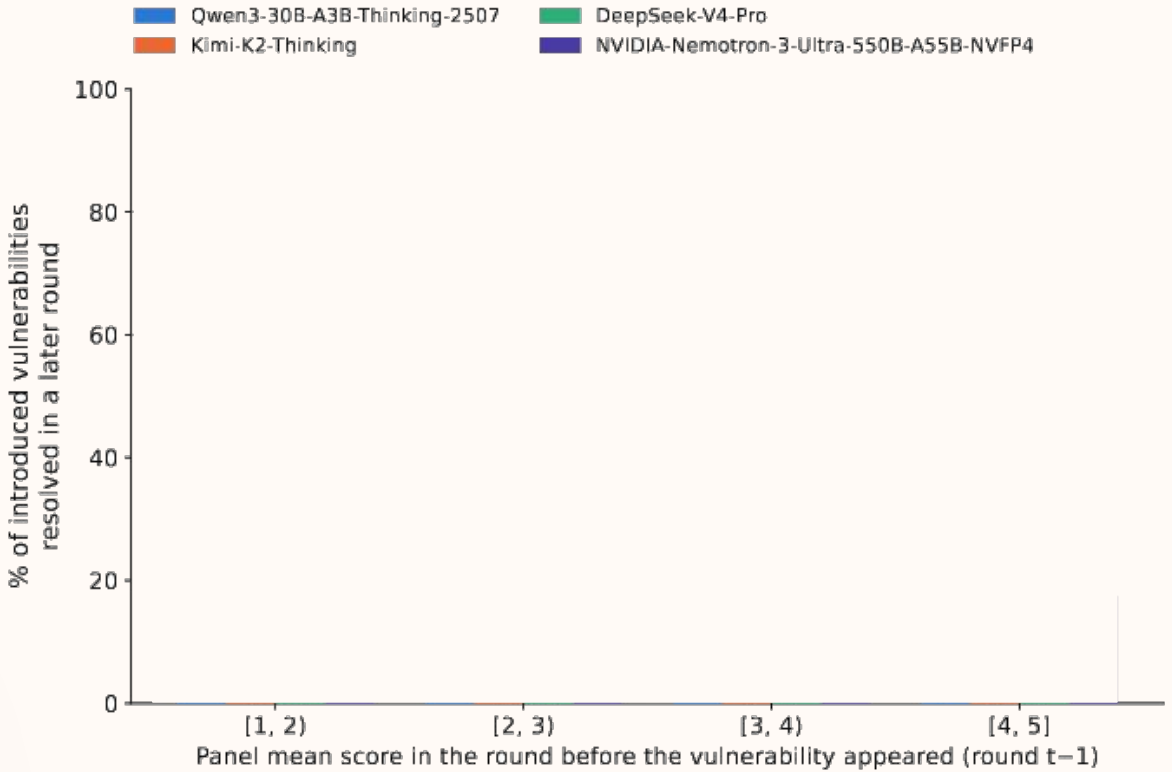
Impact of Polarization on the System



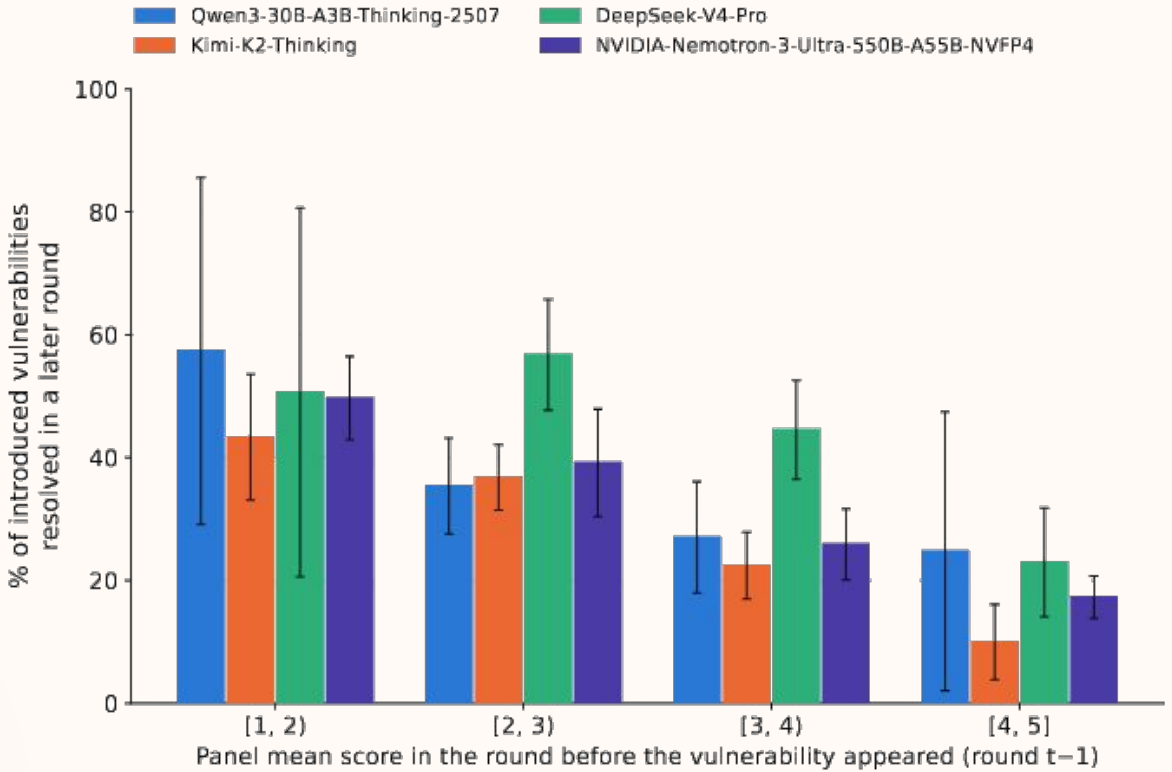
Impact of Polarization on the System



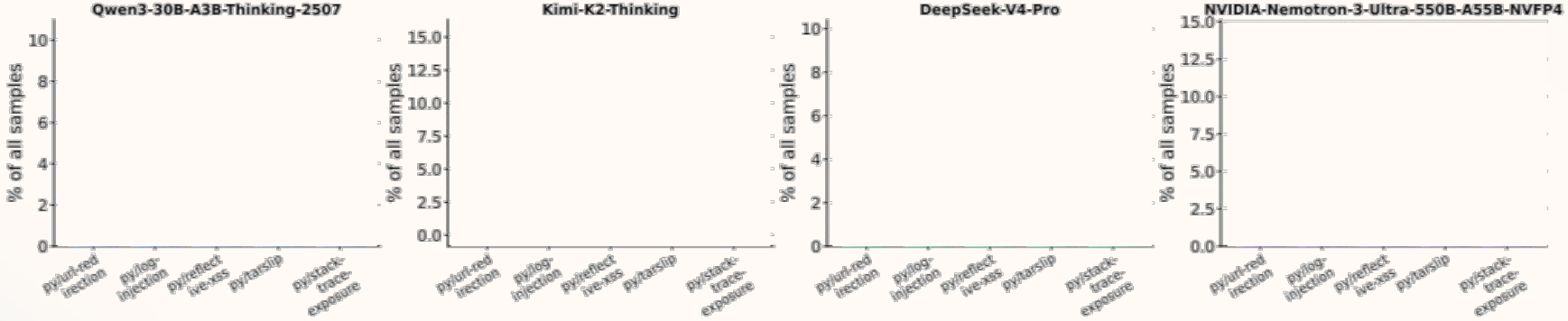
Impact of Inertia on New Vulnerabilities



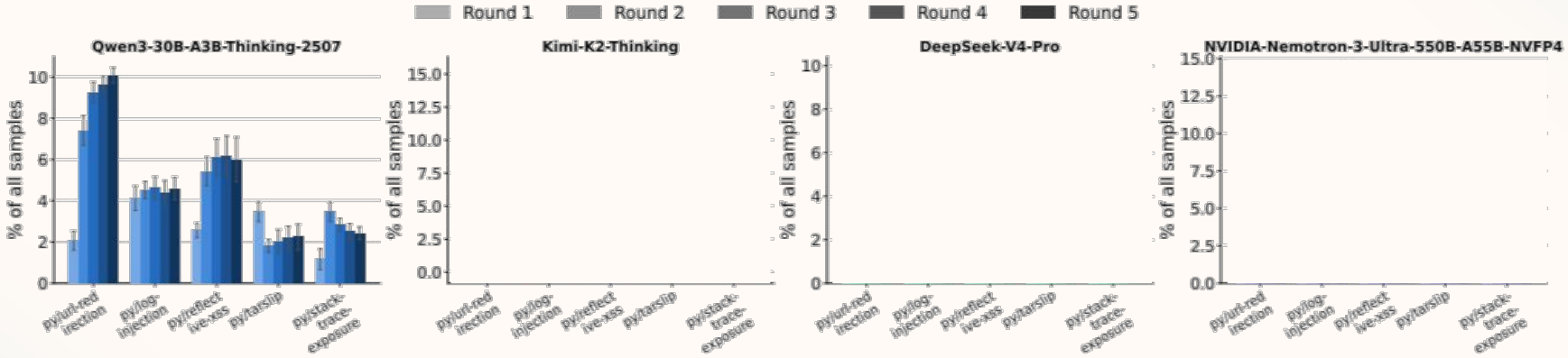
Impact of Inertia on New Vulnerabilities



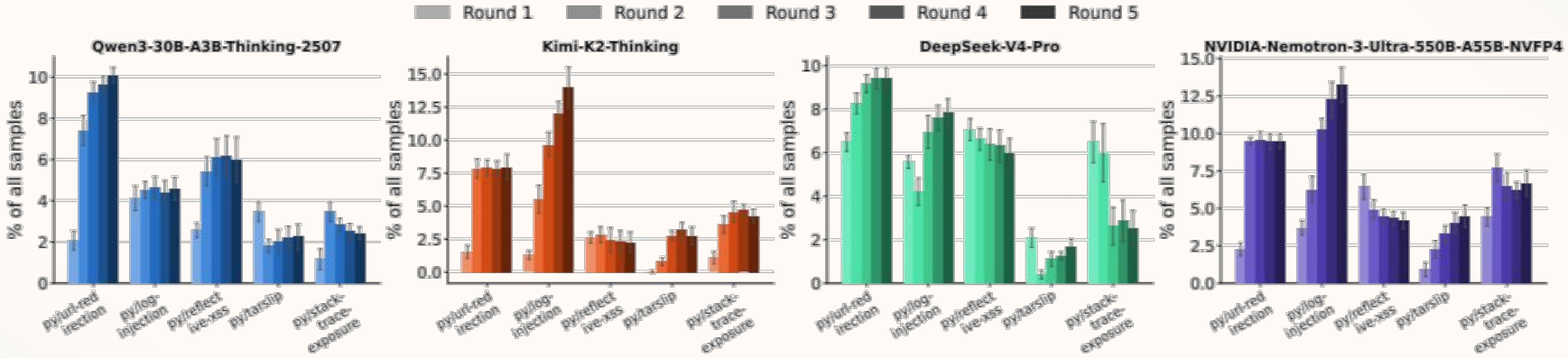
Correlated Errors in Coding



Correlated Errors in Coding



Correlated Errors in Coding



Multi-Agent Systems for Hiring

- Introducing biased prompt in a single agent spreads bias quickly to other agents due to conformity.
- Even after the biased prompt is removed, its presence in the conversation history keeps the system biased because of inertia.
- Biases once introduced not only persist but grow as the system polarizes.

Multi-Agent Systems for Hiring

- Introducing biased prompt in a single agent spreads bias quickly to other agents due to conformity.
- Even after the biased prompt is removed, its presence in the conversation history keeps the system biased because of inertia.
- Biases once introduced not only persist but grow as the system polarizes.

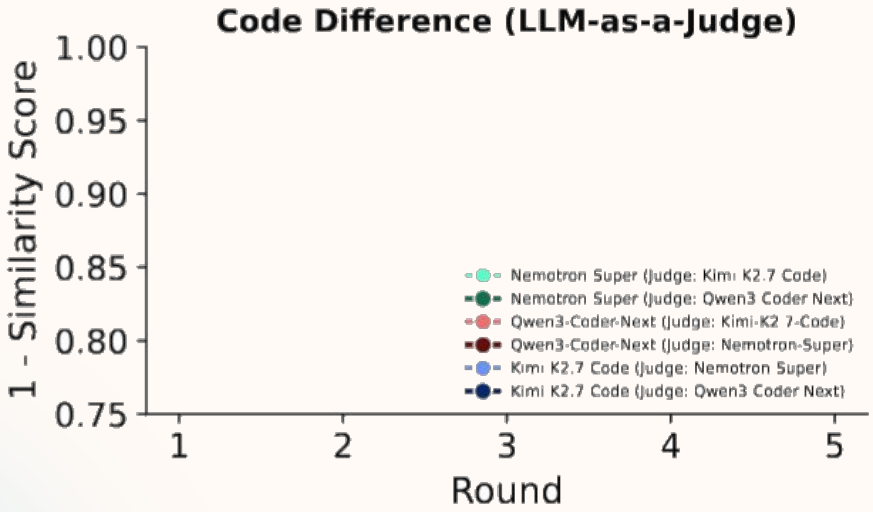
Multi-Agent Systems for Peer Review

- There is divergence in scores across different research fields, creating a disparity driven by conformity and polarization, and locked in by inertia.
- The same research fields (e.g., learning theory, reinforcement learning) benefit from it, while the same fields are affected negatively (e.g., infrastructure, software libraries, transfer learning, meta learning) across all LLMs.

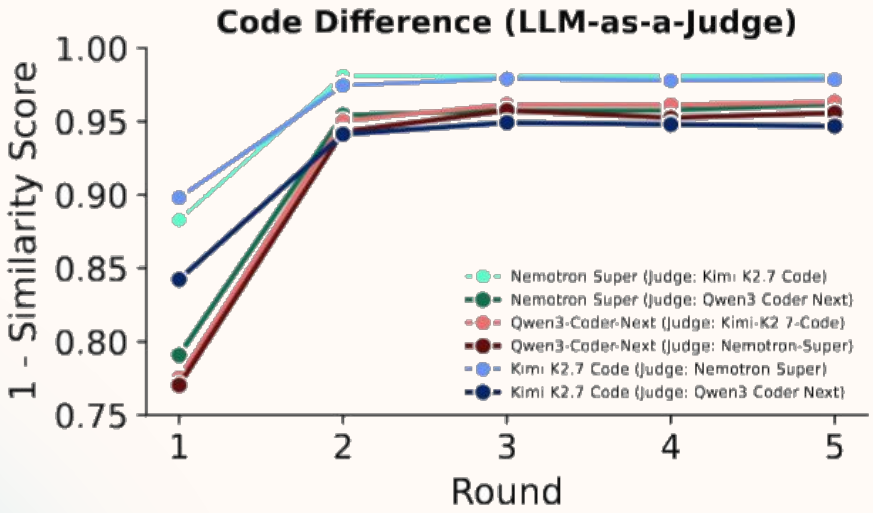
Surface-level Diversity Doesn’t Help

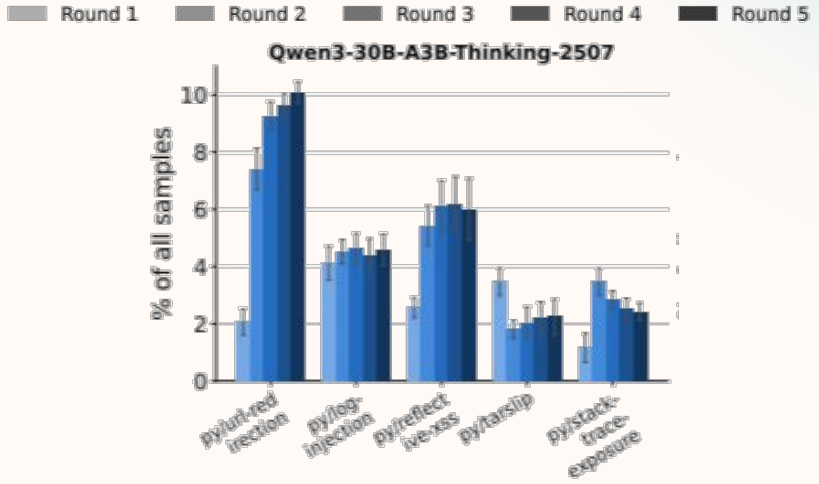


Sampling Stochasticity



Sampling Stochasticity





Sampling Stochasticity

Is Temperature the Creativity Parameter of Large Language Models?

Max Peeperkorn,[1] Tom Kouwenhoven,[2] Dan Brown,[3] and Anna Jordanous[1]

[1]School of Computing, University of Kent, United Kingdom
[2]Leiden Institute of Advanced Computer Science, Universiteit Leiden, Netherlands
[3]Cheriton School of Computer Science, University of Waterloo, Canada

m.peeperkorn@kent.ac.uk, t.kouwenhoven@liacs.leidenuniv.nl, dan.brown@uwaterloo.ca, a.k.jordanous@kent.ac.uk

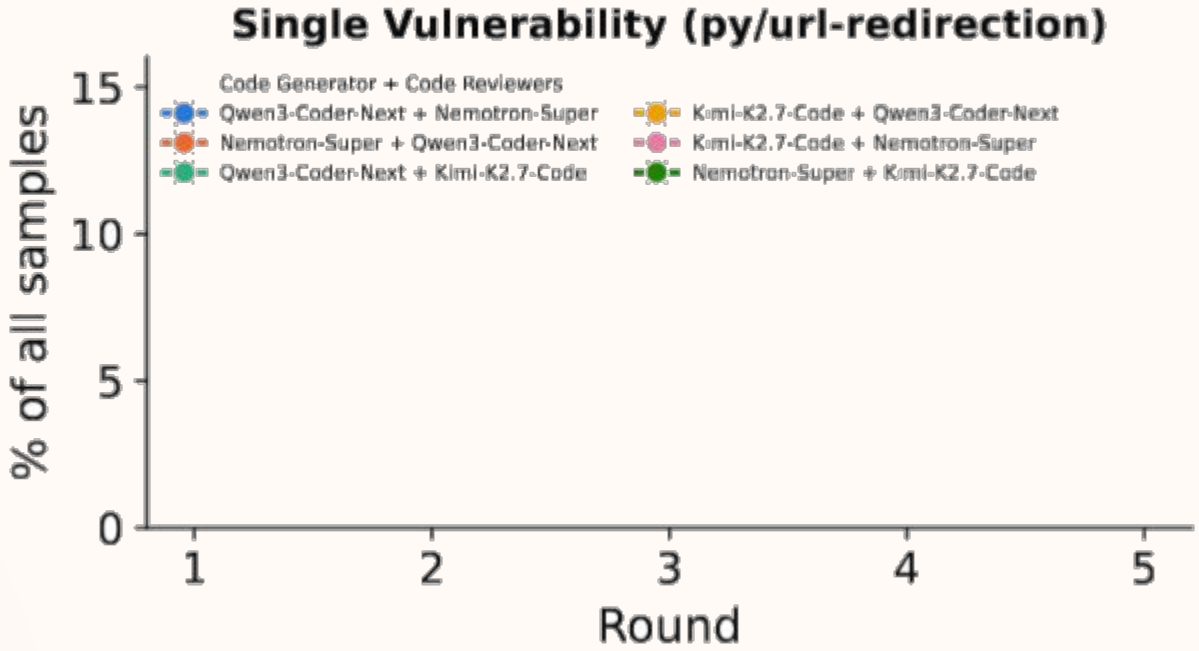
Benchmarking Linguistic Diversity of Large Language Models

Yanzhu Guo*
ALMAnaCH
Inria Paris, France
yanzhu.guo@inria.fr

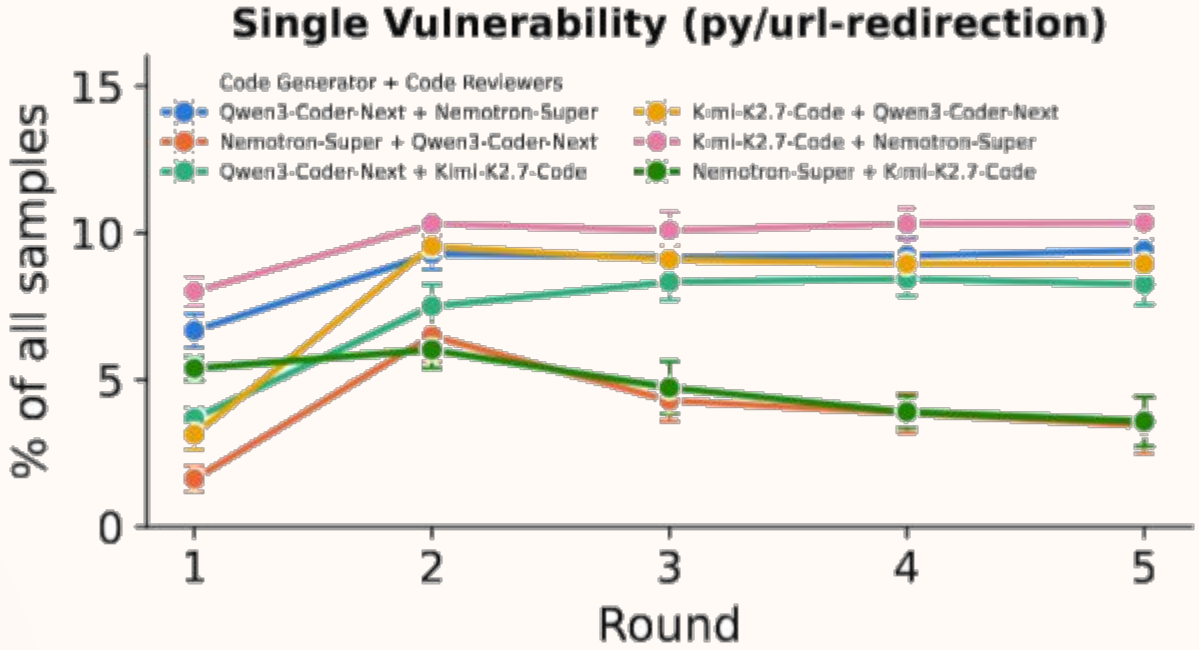
Guokan Shang
IFM Paris
MBZUAI, France
guokan.shang@mbzuai.ac.ae

Chloé Clavel
ALMAnaCH
Inria Paris, France
chloe.clavel@inria.fr

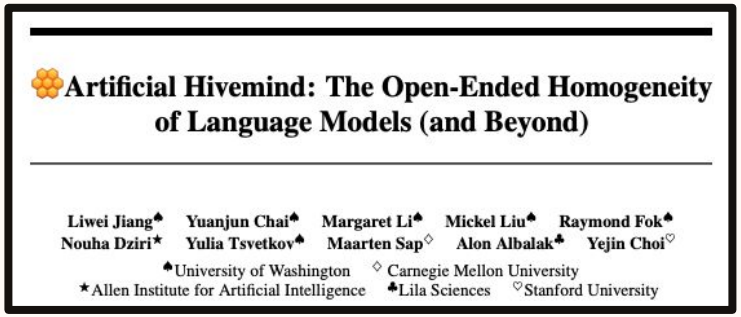
Mixed Configurations



Mixed Configurations



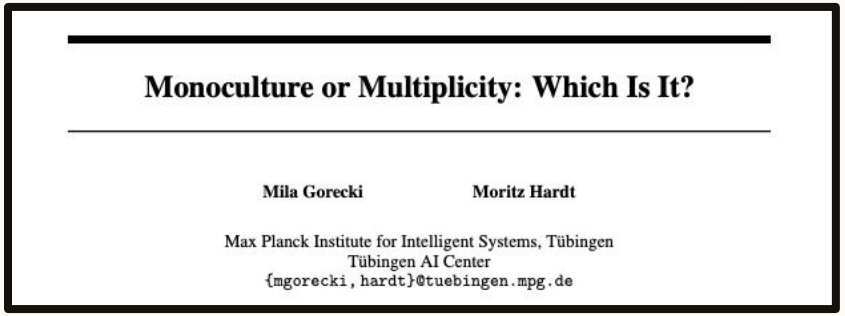
Mixed Configurations



Artificial Hivemind: The Open-Ended Homogeneity of Language Models (and Beyond)

Liwei Jiang♠ Yuanjun Chai♠ Margaret Li♠ Mickel Liu♠ Raymond Fok♠
Nouha Dziri★ Yulia Tsvetkov♠ Maarten Sap◇ Alon Albalak♣ Yejin Choi♡

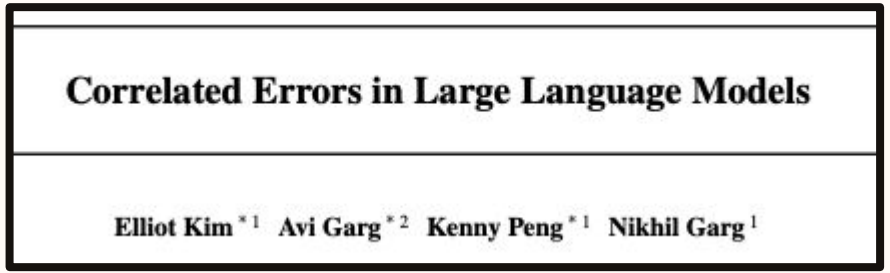
♠University of Washington ◇ Carnegie Mellon University
★Allen Institute for Artificial Intelligence ♣Lila Sciences ♡Stanford University



Monoculture or Multiplicity: Which Is It?

Mila Gorecki Moritz Hardt

Max Planck Institute for Intelligent Systems, Tübingen
Tübingen AI Center
{mgorecki, hardt}@tuebingen.mpg.de



Correlated Errors in Large Language Models

Elliot Kim [*1] Avi Garg [*2] Kenny Peng [*1] Nikhil Garg [1]

TL;DR



Homogenization emerges as a novel failure mode in Multi-Agent Systems, resulting in correlated errors, persistent perturbations, and converging preferences.

Homogenization can lead to significant downstream harms, and is not mitigated by simply injecting surface-level diversity.

Homogenization emerges as a novel failure mode in Multi-Agent Systems, resulting in correlated errors, persistent perturbations, and converging preferences.

Homogenization can lead to significant downstream harms, and is not mitigated by simply injecting surface-level diversity.

Paper (on ArXiv)

Homogenization emerges as a novel failure mode in Multi-Agent Systems, resulting in correlated errors, persistent perturbations, and converging preferences.
Homogenization can lead to significant downstream harms, and is not mitigated by simply injecting surface-level diversity.



Paper (on ArXiv)